Genomics Data 14 (2017) 132–140

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC)

Lushun Yuan[a], Liang Chen[a], Kaiyu Qian[a,b], Guofeng Qian[c], Chin-Lee Wu[d], Xinghuan Wang[a,*], Yu Xiao[a,e,f,**]

[a] Department of Urology, Zhongnan Hospital of Wuhan University, Wuhan, China
[b] Department of Urology, The Fifth Hospital of Wuhan, Wuhan, China
[c] Department of Endocrinology, The First Affiliated Hospital of Zhejiang University, Hangzhou, China
[d] Department of Urology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[e] Laboratory of Precision Medicine, Zhongnan Hospital of Wuhan University, Wuhan, China
[f] Department of Biological Repositories, Zhongnan Hospital of Wuhan University, Wuhan, China

ARTICLE INFO

Keywords:
Clear cell renal cell carcinoma (ccRCC)
Co-expression network analysis
Hub genes
Progression
Prognosis

ABSTRACT

Human clear cell renal cell carcinoma (ccRCC) is one of the most common types of malignant adult kidney tumors. We constructed a weighted gene co-expression network to identify gene modules associated with clinical features of ccRCC (n = 97). Six hub genes (*CCNB2*, *CDC20*, *CEP55*, *KIF20A*, *TOP2A* and *UBE2C*) were identified in both co-expression and protein-protein interaction (PPI) networks, which were highly correlated with pathologic stage. The significance of expression of the hub genes in ccRCC was ranked top 4 among all cancers and correlated with poor prognosis. Functional analysis revealed that the hub genes were significantly enriched in cell cycle regulation and cell division. Gene set enrichment analysis suggested that the samples with highly expressed hub gene were correlated with cell cycle and p53 signaling pathway. Taken together, six hub genes were identified to be associated with progression and prognosis of ccRCC, and they might lead to poor prognosis by regulating p53 signaling pathway.

1. Introduction

Renal cell carcinoma (RCC) is the most common type of malignant adult kidney tumors, accounting for > 90% of all adult renal tumors. Up to one-third of patients with RCC already suffered with a distant metastasis at the time of diagnosis [1]. Clear cell RCC (ccRCC), taking up about 75%–85% of RCC, is the most common subtype [2]. At present, ccRCC is usually resistant to chemotherapy. Targeted therapies have been exploited for their target specificity and low toxicity so they can be the best choice of non-surgical treatments [3]. Therefore, many biomarkers for clear cell renal cell carcinoma have been discovered including *VHL*, *VEGF*, *CAIX* and *HIF1α/2α* mutations. Some of which could predict therapeutic effect and clinical prognosis [4]. We know that carcinogenesis is not the result of deregulation of several oncogenes or tumor suppressors; it is the outcome of complex mechanisms, including the high interconnection between genes with similar expression patterns [5]. Thus, it is urgently needed to identify novel molecular biomarkers that can predict disease stage and clinical outcome of ccRCC patients, which could help understand its pathogenesis and provide personalized treatment.

Rapid technological breakthroughs of genome-wide sequencing have shed new light on the research of clinical issues and related pathological mechanisms in various cancers [6]. Nowadays, most studies just concentrated on the screening of differentially expressed genes and not attached enough attention to the high degree of interconnection between genes, where genes with similar expression patterns may be functionally related. The algorithm, weighted gene co-expression network analysis (WGCNA), can construct free-scale gene co-expression networks to explore the relationships between different gene sets or between gene sets and clinical features [7]. WGCNA has been widely

Abbreviations: ccRCC, clear cell renal cell carcinoma; HPA, human protein atlas; GSEA, enrichment analysis and gene set enrichment; WGCNA, weighted gene co-expression network analysis; TCGA, the cancer genome atlas; DEGs, differentially expressed genes; TOM, topological overlap matrix; MEs, module eigengenes; GS, gene significance; MS, module significance; STRING, search tool for the retrieval of interacting genes; PPI, protein-protein interaction; DAVID, Database for Annotation, Visualization and Integrated Discovery; DEG, differentially expressed gene; SAM, significance analysis of microarrays

* Corresponding author.
** Correspondence to: Y. Xiao, Department of Biological Repositories, Zhongnan Hospital of Wuhan University, Wuhan, China.
E-mail addresses: wangxinghuan@whu.edu.cn (X. Wang), yu.xiao@whu.edu.cn (Y. Xiao).

https://doi.org/10.1016/j.gdata.2017.10.006
Received 10 July 2017; Received in revised form 12 October 2017; Accepted 25 October 2017
Available online 04 November 2017

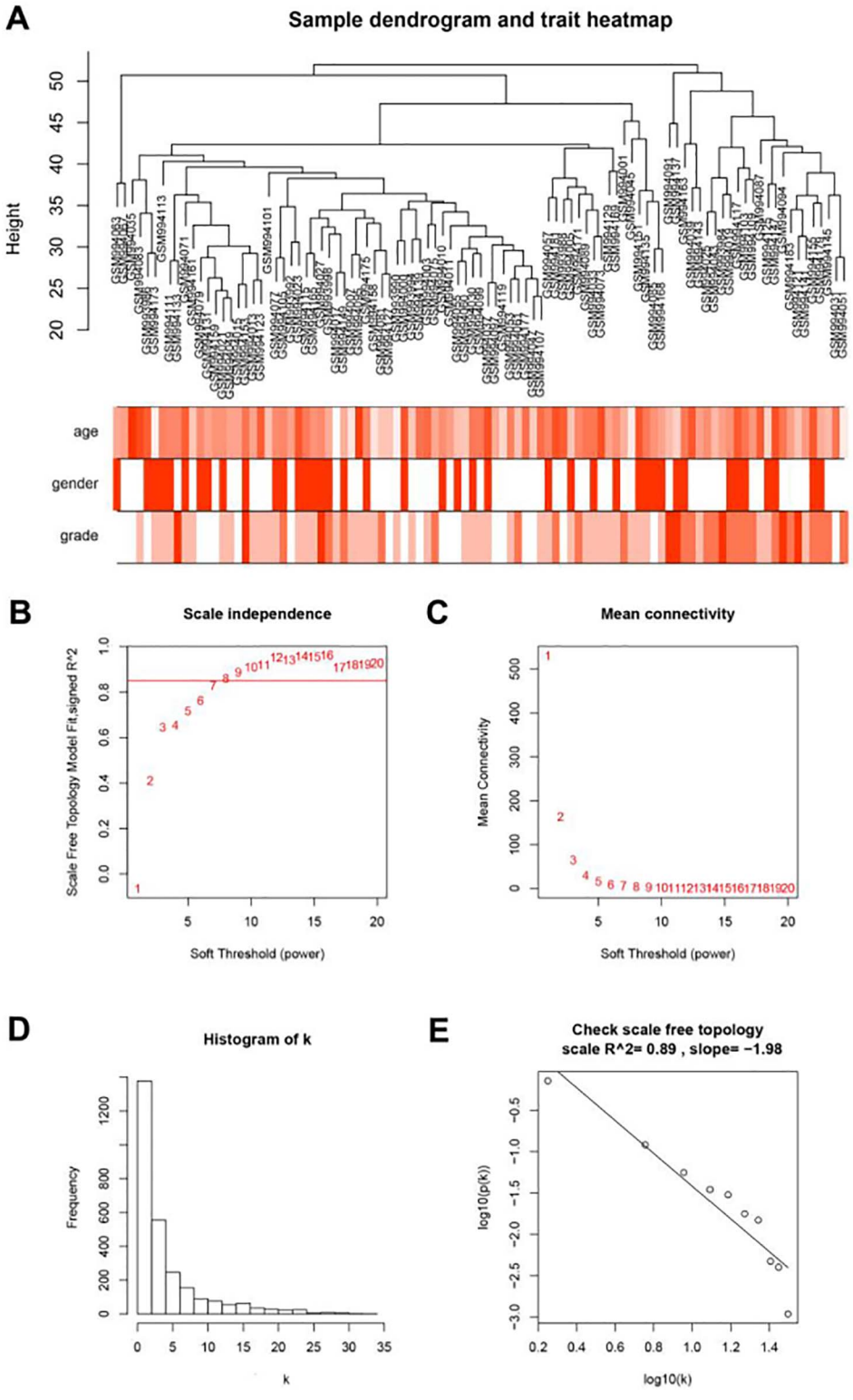

Fig. 1. Clustering dendrogram of 97 tumor samples and the clinical traits. (A) The clustering was based on the expression data of differentially expressed genes between tumor samples and non-tumor samples in ccRCC. The color intensity was proportional to older age and higher Furhman grade. (B-D) Determination of soft-thresholding power in the weighted gene co-expression network analysis (WGCNA). (B) Analysis of the scale-free fit index for various soft-thresholding powers (β). (C) Analysis of the mean connectivity for various soft-thresholding powers. (D) Histogram of connectivity distribution when β = 8. (E) Checking the scale free topology when β = 8. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

applied to finding the hub genes associated with clinical feature in different cancer types [8–10].

In this study, WGCNA and other analysis methods are adopted to jointly analyze clinical information and microarray data of ccRCC patient samples to identify key genes associated with clinical features (age, gender, and tumor grade). These key genes may have important clinical implications and serve as diagnostic and prognostic biomarkers or therapeutic targets.

2. Materials and methods

2.1. Data collection

Expression profiles of mRNA and related clinical data of clear cell renal carcinoma were downloaded from Gene Expression Omnibus (GEO) database (*http://www.ncbi.nlm.nih.gov/geo/*). Dataset GSE40435 performed on Illumina HumanHT-12 V4.0 expression beadchip was used as a training set to construct co-expression networks and identify

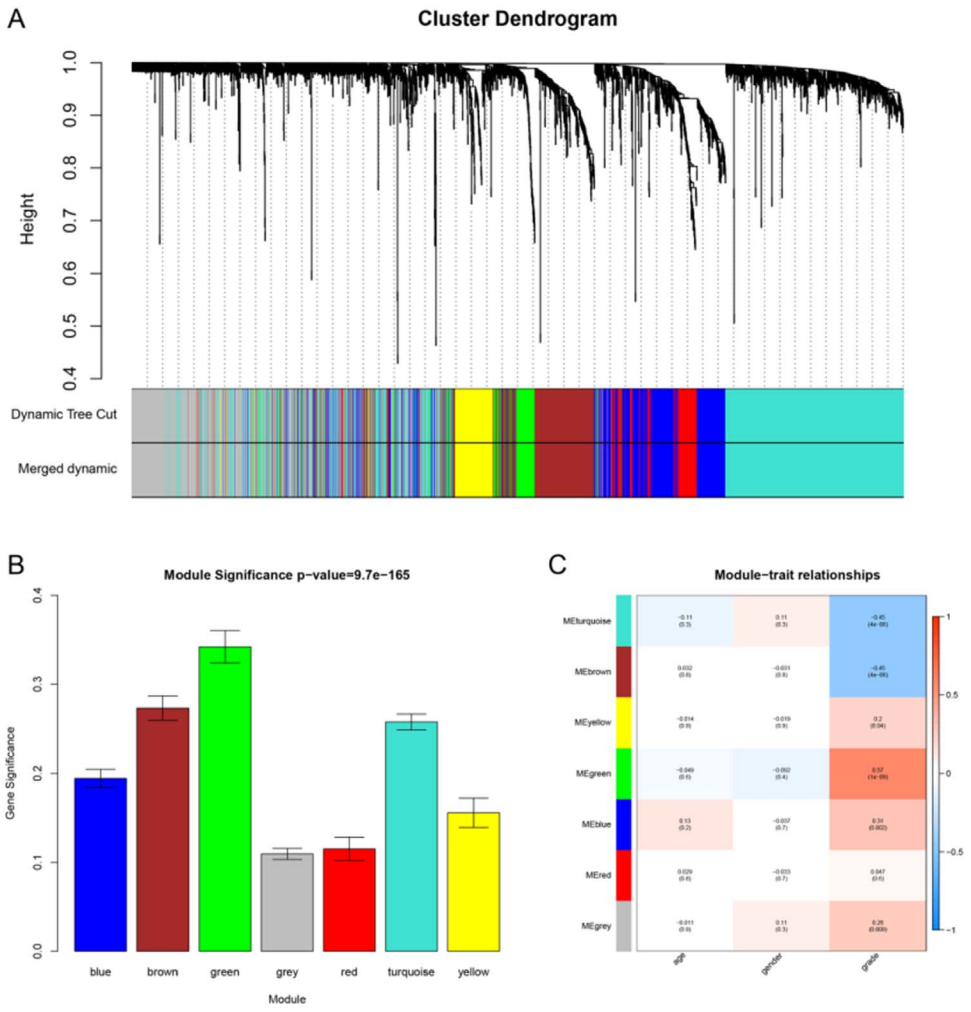

Fig. 2. Identification of modules associated with clinical information. (A) Dendrogram of all differentially expressed genes clustered based on a dissimilarity measure (1-TOM). (B) Distribution of average gene significance and errors in the modules associated with the Furhman grade of ccRCC. (C) Heatmap of the correlation between module eigengenes and different clinical information of ccRCC (age, gender, and Furhman grade).

hub genes in this study. This dataset included 101 pairs of ccRCC tumors and adjacent non-tumor renal tissue from Czech patients (including 22 ccRCC of grade 1, 47 of grade 2, 24 of grade 3, and 8 of grade 4). Another independent dataset of GSE53757 based on the microarray platform of Affymetrix U133 plus 2 was downloaded from GEO and used as a test set to verify our results. This dataset included 72 pairs of clear cell renal carcinoma patients (including ccRCC of grade 1, 2, 3, and 4).

2.2. Data preprocessing

Normalized data were firstly downloaded from GEO database. Microarray quality was assessed by sample clustering according to the distance between different samples in Pearson's correlation matrices and average linkage, and 4 samples (GSM993996, GSM994041, GSM994065, and GSM994069) were removed from subsequent analysis in GSE40435.

2.3. Differentially expressed genes (DEGs) screening

The “limma” (linear models for microarray data) R package was used to screen the DEGs between normal kidney and clear cell renal cell carcinoma. The SAM (significance analysis of microarrays) with FDR (false discovery rate) < 0.05 and |log2 fold change (FC)| > 0.585 were chosen as the cut-off criteria to select genes further considered in the network construction.

2.4. Co-expression network construction

Firstly, expression data profile of DEGs was tested to check if they were the good samples and good genes. Then, we used the “WGCNA” package in R to construct scale-free co-expression network for the DEGs. At first, the Pearson's correlation matrices and average linkage method were both performed for all pair-wise genes. Then, a weighted adjacency matrix was constructed using a power function $a_{mn} = |c_{mn}|^{\beta}$ (c_{mn} = Pearson's correlation between gene m and gene n; a_{mn} = adjacency between gene m and gene n). β was a soft-thresholding parameter that could emphasize strong correlations between genes and penalize weak correlations. After choosing the power of β, the adjacency was transformed into a topological overlap matrix (TOM), which could measure the network connectivity of a gene defined as the sum of its adjacency with all other genes for network generation, and the corresponding dissimilarity (1-TOM) was calculated. To classify genes with similar expression profiles into gene modules, average linkage hierarchical clustering was conducted according to the TOM-based dissimilarity measure with a minimum size (gene group) of 50 for the genes dendrogram. To further analyze the module, we calculated the dissimilarity of module eigengenes, chose a cut line for module dendrogram and merged some module.

2.5. Identification of clinically significant modules

Two approaches were used to identify modules related to clinical traits of ccRCC. First, gene significance (GS) was defined as the log10 transformation of the *P* value (GS = lgP) in the linear regression

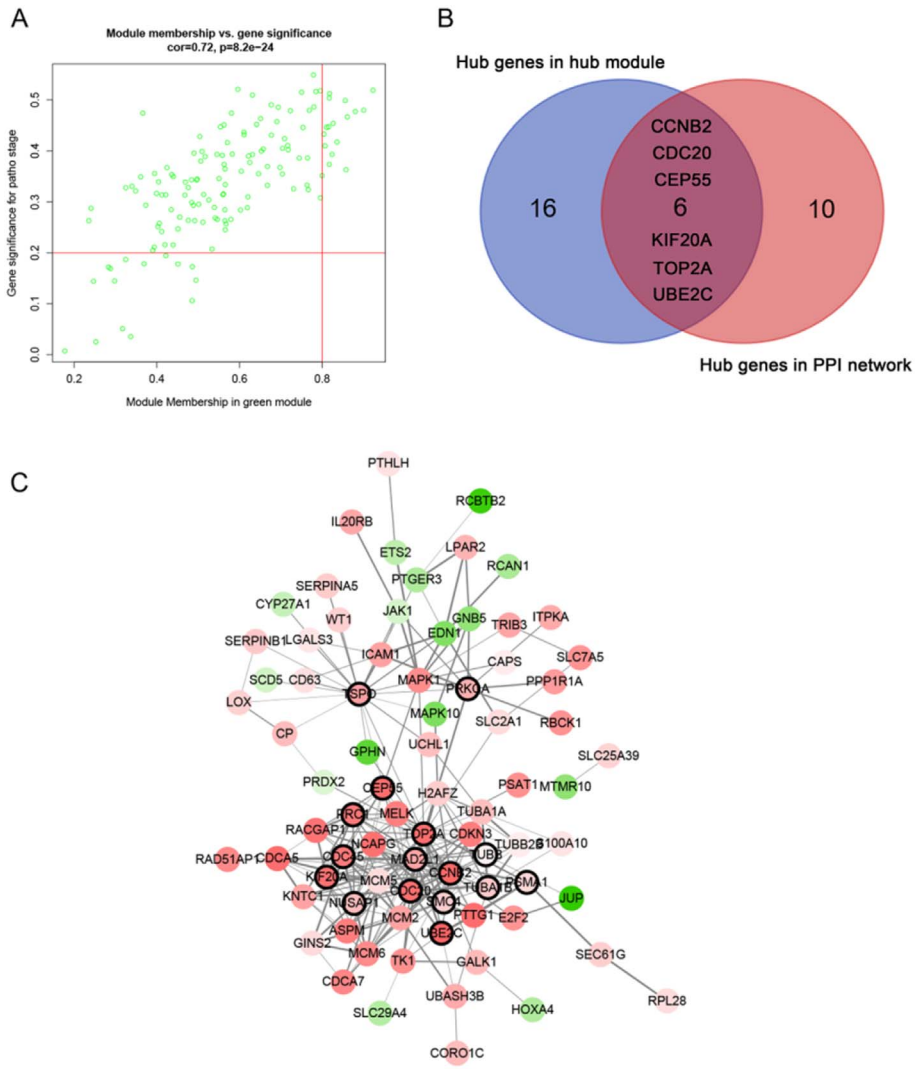

Fig. 3. Hub genes detection and protein-protein network (PPI). (A) Scatter plot of module eigengenes in green module. (B) The Venn diagram of co-expression hub genes and PPI network hub genes. (C) Protein–protein interaction network of genes in the green module. The color intensity in each node was proportional to fold change of expression in comparison to normal kidney samples (up-regulation in red and down-regulation in green). The nodes with bold circle represented network hub genes identified by WGCNA. The edge width was proportional to the score of protein-protein interaction based on the STRING database. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between gene expression and the clinical traits. In addition, module significance (MS) was defined as the average GS for all the genes in a module. In general, the module with the absolute MS ranked first or second among all the selected modules was considered as the one related with clinical trait. Module eigengenes (MEs) were considered as the major component in the principal component analysis for each gene module, and the expression patterns of all genes could be summarized into a single characteristic expression profile within a given module. In addition, we calculated the correlation between MEs and clinical traits to identify the relevant module. The module with the maximal absolute MS among all the selected modules was usually considered as the one related with clinical trait. Finally, the module highly correlated with certain clinical trait was selected for further analysis.

2.6. Identification of hub genes

Hub genes, highly interconnected with nodes in a module, have been shown to be functionally significant. In this study, hub genes were defined by module connectivity, measured by absolute value of the Pearson's correlation (cor.geneModuleMembership > 0.8) and clinical trait relationship, measured by absolute value of the Pearson's correlation (cor.geneTraitSignificance > 0.2). We identified hub genes in module which were highly correlated with certain clinical trait. Furthermore, we uploaded all genes in the hub module to the STRING database, choosing confidence > 0.4 to construct protein-protein interaction (PPI). In the PPI network, genes with a connectivity degree of ≥ 8 were also defined hub genes. The common hub genes both in co-expression network and PPI network were regarded as “real” hub genes for further analysis.

2.7. Hub genes validation

In the test set of GSE53757, linear regression analyses were performed to validate the role of hub genes in the progression of ccRCC as well as the transcriptional levels in normal kidney and ccRCC samples. Moreover, we used 2 other databases: Gene Expression Profiling Interactive Analysis (*http://gepia.cancer-pku.cn*) and Oncolnc (*http://www.oncolnc.org*) to perform validation of cancer specific expression and prognosis of the candidate hub genes [11,12].

2.8. Functional and pathway enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (*http://david.abcc.ncifcrf.gov/*) is an online program providing a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes [13]. We uploaded DEGs in hub module to perform gene ontology and KEGG pathway enrichment analysis. Those terms including hub genes were selected as the key biological process and pathways. $P < 0.05$ was set as the cut-off criterion.

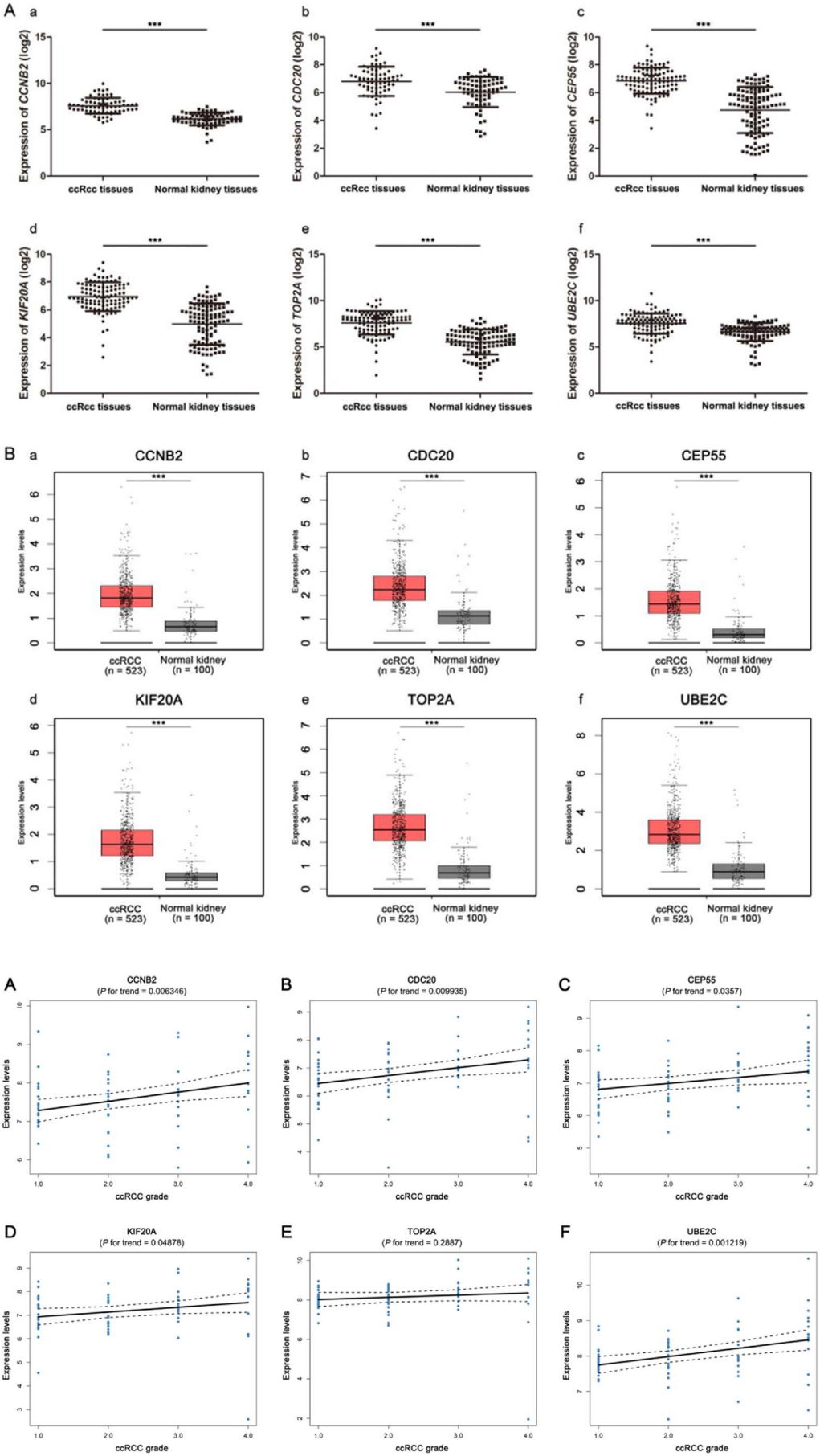

Fig. 4. Validation of the gene expression levels of *CCNB2*, *CDC20*, *CEP55*, *KIF20A*, *TOP2A* and *UBE2C* between normal kidney and ccRCC samples. (A) Validation based on microarray data of GSE53757. (a) *CCNB2*, (b) *CDC20*, (c) *CEP55*, (d) *KIF20A*, (e) *TOP2A*, and (f) *UBE2C*. (B) Validation based on TCGA data in GEPIA. (a) *CCNB2*, (b) *CDC20*, (c) *CEP55*, (d) *KIF20A*, (e) *TOP2A*, and (f) *UBE2C*.

Fig. 5. Validation of the correlation between the expression levels of *CCNB2*, *CDC20*, *CEP55*, *KIF20A*, *TOP2A*, and *UBE2C* and the Furhman grade of ccRCC (based on microarray data of GSE53757). (A) *CCNB2*, (B) *CDC20*, (C) *CEP55*, (D) *KIF20A*, (E) *TOP2A*, and (F) *UBE2C*.

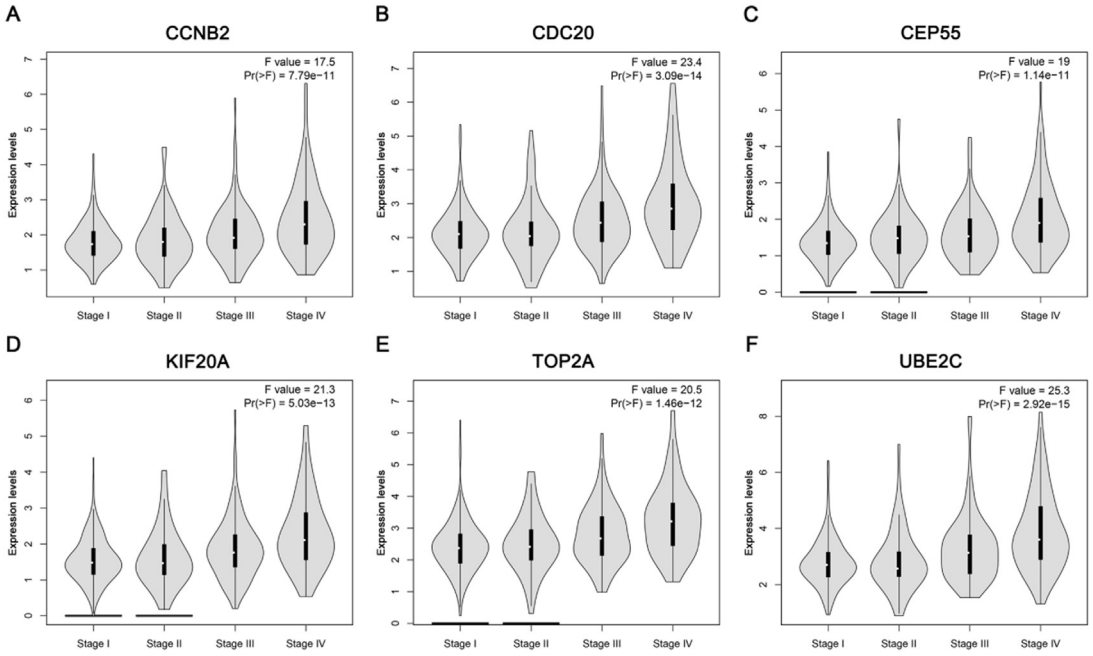

Fig. 6. Validation of the correlation between the expression levels of *CCNB2*, *CDC20*, *CEP55*, *KIF20A*, *TOP2A* and *UBE2C* and the pathologic stage of ccRCC (based on TCGA data in GEPIA). (A) *CCNB2*, (B) *CDC20*, (C) *CEP55*, (D) *KIF20A*, (E) *TOP2A*, and (F) *UBE2C*.

2.9. Gene set enrichment analysis (GSEA)

In the test set, ccRCC we firstly seek out the median expression of each hub gene and according to the median, 72 ccRCC samples were divided into two groups. To identify potential function of the hub gene, GSEA (*http://software.broadinstitute.org/gsea/index.jsp*) was conducted to detect whether a series of a priori defined biological processes were enriched in the gene rank derived from DEGs between the two groups [14,15]. For use with GSEA software, the collection of annotated gene sets of c2.cp.kegg.v6.0.symbols.gmt in Molecular Signatures Database (*MSigDB*, *http://software.broadinstitute.org/gsea/msigdb/index.jsp*) was chosen as the reference gene sets. FDR < 0.05 was chosen as the cut-off criterion.

2.10. Analysis of DEGs screening between grade I-III and grade IV

One hundred one samples were divided into 2 groups (grade I-III and grade IV). The FDR < 0.05 and $|\log2$ fold change (FC)$| > 0.585$ were chosen as DEGs between low grade ccRCC and high grade ccRCC. Then DEGs were uploaded to DAVID database to perform functional enrichment analysis.

3. Results

3.1. DEGs screening

After data preprocessing and quality assessment, the expression matrices were obtained from the 202 samples in training set GSE40435. Under the threshold of FDR < 0.05 and $|\log_2 FC| > 0.585$, a total of 2756 DEGs (1314 up-regulated and 1442 down-regulated) were selected for subsequent analysis.

3.2. Weighted co-expression network construction and key modules identification

Ninety-four over one hundred one samples with clinical data were included in co-expression analysis (Fig. 1A). Using "WGCNA" package in R, the DEGs with similar expression patterns were grouped into modules via the average linkage hierarchical clustering. In this study, the power of $\beta = 8$ (scale free $R^2 = 0.89$) was selected as the soft-thresholding to ensure a scale-free network (Fig. 1B-E). A total of 7 modules were identified (Fig. 2A). Two methods were used to test the relevance between each module and the ccRCC progression. Firstly, modules with greater MS were considered to have more connection with the disease progression. However, most of the correlations were low to moderate ($R^2 < 0.5$), and we found that the MS of green module ($P = 1 \times 10^{-9}$, $R^2 = 0.57$) was higher than those of any other MS (Fig. 2B). Afterwards, the ME in the green module showed a higher correlation with disease progression than other modules (Fig. 2C). Based on the two methods, the green module with tumor progression was identified as the clinical significant module, which was extracted for further analysis.

3.3. Identification of hub genes for tumor progression in the green module

Highly connected hub genes in a module play important roles in the biological processes. Defined by module connectivity, measured by absolute value of the Pearson's correlation (cor.geneModuleMembership > 0.8) and clinical trait relationship, measured by absolute value of the Pearson's correlation (cor.geneTraitSignificance > 0.2), 22 genes with the high connectivity in green module were taken as hub genes (*TOP2A*, *UHRF1*, *RAD51AP1*, *NR3C2*, *MCM6*, *PTTG1*, *NCAPG*, *CDCA5*, *UBE2C*, *PDGFD*, *RGS5*, *CEP55*, *CCNB2*, *CDKN3*, *CCDC109B*, *PTTG3P*, *RACGAP1*, *KIF20A*, *ALDH3A2*, *CDC20*, *EDNRB*, and *EMCN*) (Fig. 3A). As to the PPI network, under the cutoff of confidence > 0.4 and connectivity degree of ≥ 8, 16 genes were identified as hub genes (*CCNB2*, *CDC20*, *CDC45*, *CEP55*, *KIF20A*, *MAD2L1*, *NUSAP1*, *PRC1*, *PRKCA*, *PSMA1*, *SMC4*, *TOP2A*, *TSPO*, *TUBA1B*, *TUBB*, and *UBE2C*) (Fig. 3C). *CCNB2*, *CDC20*, *CEP55*, *KIF20A*, *TOP2A*, and *UBE2C* were identified both in PPI network and co-expression network (Fig. 3B). Eventually, these six genes were regarded as "real" hub genes for tumor progression, which were chosen for further analysis.

3.4. Hub gene validation

Gene expression validations were performed, and all of six hub genes were also oncogenes in the test set and TCGA data (Fig. 4). Then, linear regression analyses were conducted to validate hub genes in the test set GSE53757, most of which except *TOP2A* showed positive

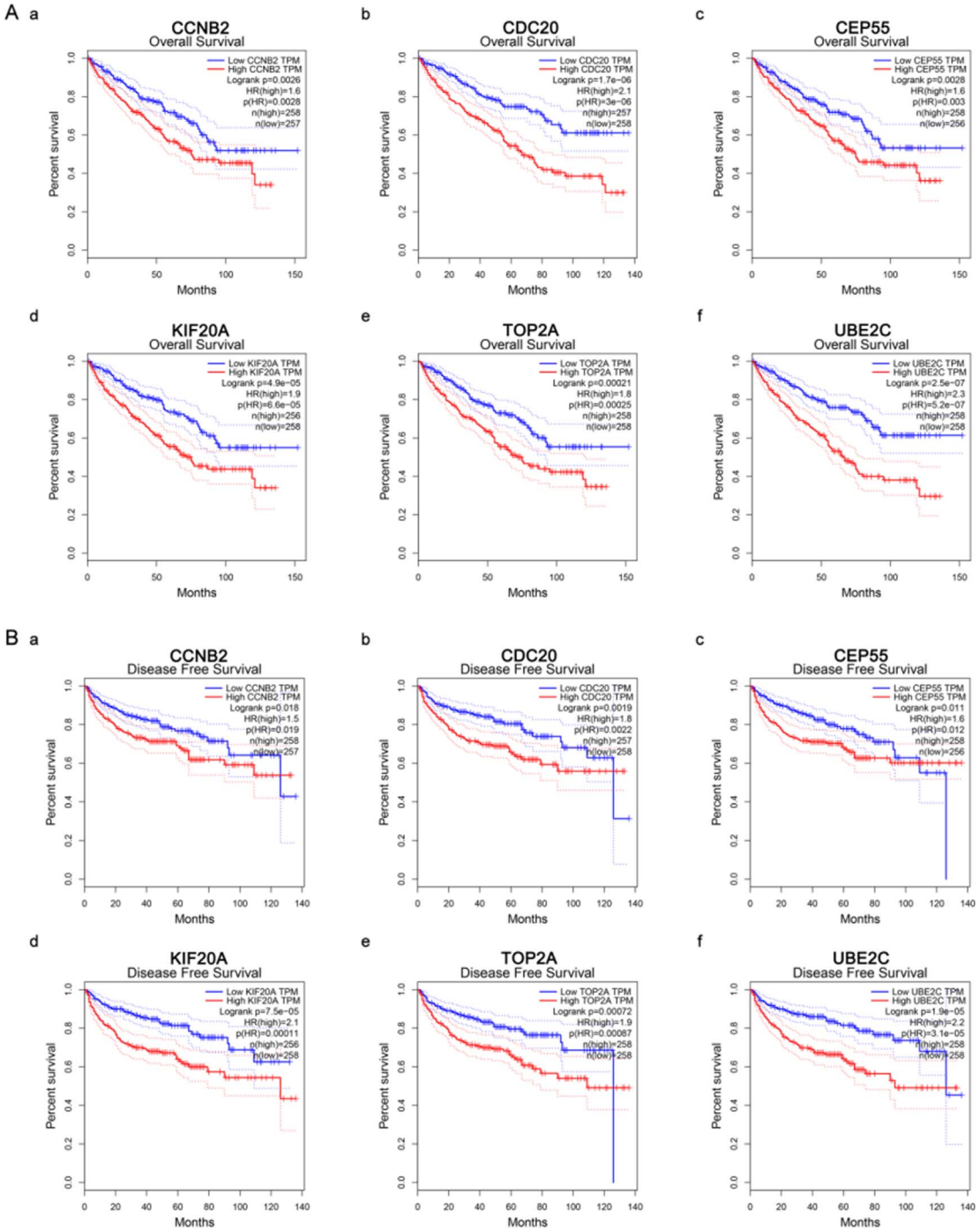

Fig. 7. Survival analysis of six hub genes in ccRCC (based on TCGA data in GEPIA). (A) Overall survival analysis. (a) *CCNB2*, (b) *CDC20*, (c) *CEP55*, (d) *KIF20A*, (e) *TOP2A*, and (f) *UBE2C*. (B) Disease free survival analysis. (a) *CCNB2*, (b) *CDC20*, (c) *CEP55*, (d) *KIF20A*, (e) *TOP2A*, and (f) *UBE2C*. Red line represented the samples with gene highly expressed and blue line was for the samples with gene lowly expressed. HR: hazard ratio. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

correlation with ccRCC progression (Furhman grade) (P for trend < 0.05) (Fig. 5). Moreover, based on the TCGA data, we investigated that all of six genes were highly-correlated with pathologic stage (Fig. 6). As the tumor progression always affected the tumor prognosis, we also validated the six hub genes by investigating their roles in ccRCC prognosis including overall survival time and disease free survival time. We found a shorter overall survival time (Fig. 7A) and disease-free survival time in patients with higher expression levels of *CCNB2*, *CDC20*, *CEP55*, *KIF20A*, *TOP2A*, and *UBE2C* (Fig. 7B). In addition, based on the Oncolnc database, we found a cancer-specific expression of the six hub genes, which demonstrated that the significance of expression of *CCNB2*, *CDC20*, *CEP55*, *KIF20A, TOP2A*, and *UBE2C* was top 4 among all cancers (Suplplementary Fig. S1).

3.5. Functional and pathway enrichment analysis

To obtain further insight into the function of DEGs in hub module, they were uploaded to the DAVID database. GO analysis results showed that hub genes were enriched in the top 10 biological process (BP), including proteasome-mediated ubiquitin-dependent protein catabolic process, negative regulation of blood coagulation, anaphase-promoting complex-dependent catabolic process, DNA unwinding involved in DNA

replication, microtubule-based process, DNA replication initiation, mitotic cytokinesis, G1/S transition of mitotic cell cycle, mitotic chromosome condensation, and cell division. Moreover, hub genes were also overrepresented in these top 10 KEGG pathways, including phagosome, progesterone-mediated oocyte maturation, DNA replication, Hepatitis B, pancreatic cancer, pathways in cancer, HTLV-1 infection, pathogenic *Escherichia coli* infection, gap junction, and cell cycle (Supplementary Fig. S2). Among the functional and pathway enrichment analysis, cell division and cell cycle were most significantly enriched.

3.6. Gene set enrichment analysis

To identify the potential function of those hub genes in ccRCC, GSEA was conducted to search biological processes enriched in any hub gene highly-expressed samples. Six gene sets were enriched in the samples with all hub genes highly expressed, including “cell cycle”, “p53 signaling pathway”, “DNA replication”, “primary immunodeficiency”, “homologous recombination”, and “base excision repair” (Supplementary Fig. S3).

3.7. Grade-associated DEGs analysis

Firstly, sample cluster was performed (Supplementary Fig. S4A).Under the threshold of FDR < 0.05 and $|\log_2 FC| > 0.585$, a total of 48 DEGs (35 up-regulated and 13 down-regulated) were identified (Supplementary Fig. S4B and Supplementary Table 1). And they were significantly enriched in following terms listed in Supplementary Table 2.

4. Discussion

The progression and prognosis of clear cell renal cell carcinoma were quite variable. Though a number of prognostic models had been proposed, most of them were based on clinical parameters and lacked accuracy. In the era of precise medicine, better biomarkers for cancer specific prognosis and progression were urgently needed and provided more accurate clinical information that could significantly enhance decision making for patient management. Here, we used an integrated analysis to screen progression and prognosis related biomarkers.

WGCNA was performed to identify gene co-expression modules related with the progression of ccRCC. The green module was identified, and 22 hub genes were derived from the module. Furthermore, relating the results of PPI network, 16 hub nodes were identified. Therefore, *CCNB2*, *CDC20*, *CEP55*, *KIF20A*, *TOP2A*, and *UBE2C* in common networks, were real hub genes, indicating that they had high connection with clinical trait as well as vital biological processes. Few of them were identified in clear cell renal cell carcinoma.

Cyclin B2 (*CCNB2*) is a member of the cyclin family, specially the B-type cyclins. Takashima S et al. reported that stronger expression of cyclin B2 mRNA in tumor cells was an independent predictor of a poor prognosis in patients with adenocarcinoma of the lung [16]. Furthermore, Shubbar E et al. also found that *CCNB2* might function as an oncogene and could serve as a potential biomarker of unfavorable prognosis over short-term follow-up in breast cancer [17]. Cell Division Cycle 20 (*CDC20*), acted as a regulatory protein interacting with several other proteins at multiple points in the cell cycle. Many studies demonstrated that *CDC20* might be a potential biomarker for both therapeutic target and prognosis in a number of cancers [18–21]. Centrosomal protein 55 (*CEP55*), played a role in mitotic exit and cytokinesis, which was important in carcinogenesis as well. Tao J. et al. suggested that *CEP55* could suppress cellular proliferation by regulating cell cycle arrest at G2/M phase in human gastric cancer [22]. *CEP55* overexpression has been found to significantly correlate with tumor stage, aggressiveness, metastasis, and poor prognosis across multiple tumor types and therefore has been included as part of several prognostic ‘gene signatures’ for cancer [23]. Kinesin Family Member 20A (*KIF20A*) might act as a motor required for the retrograde *RAB6* regulated transport of Golgi membranes and associated vesicles along microtubules. Shi C. et al. found that *Gli2-KIF20A* axis was essential for the proliferation and growth of human hepatocellular carcinoma cells as well as a potential target for future therapeutic intervention and as an independent prognostic biomarker for hepatocellular carcinoma [24]. *KIF20A* was also reported that its over-expression correlates with HPV infection, clinical stage, tumor recurrence, lymphovascular space involvement, pelvic lymph node metastasis, and poor outcome in early-stage cervical squamous cell carcinoma patients [25]. Topoisomerase (DNA) II Alpha (*TOP2A*) encoded a DNA topoisomerase, an enzyme that controlled and altered the topologic states of DNA during transcription. Study had demonstrated that *TOP2A* identified and provided epigenetic rationale for novel combination therapeutic strategies for aggressive prostate cancer [26]. Moreover, we could find the prognostic value of *TOP2A* in nasopharyngeal carcinoma and breast cancer [27,28]. Ubiquitin-conjugating enzyme E2 C (*UBE2C*), a member of the E2 ubiquitin-conjugating enzyme family, was required for the destruction of mitotic cyclins and for cell cycle progression and may be involved in cancer progression. It was reported to be a promising biomarker in tumor progression and prognosis [29–32].

Based on many studies about six hub genes, we could find the promising values of tumor progression and prognosis. According to the tumor specific expression (Supplementary Fig. S1), we found that the *p* value of each hub gene in ccRCC ranked top4 in all kinds of cancers based on the Oncolnc database, which demonstrated that those hub genes were strongly correlated with ccRCC as well as with great diagnostic value for ccRCC.

To further study their mechanism of regulating tumorigenesis, we perform the gene ontology and KEGG pathway analysis as well as GSEA analysis. In GO analysis, we found that hub genes were significantly enriched in terms of regarding cell cycle and cell cycle regulation; meanwhile, the KEGG analysis showed that cell cycle was the most significant pathway. Interestingly, through GSEA analysis, we found that hub genes were commonly enriched in cell cycle and p53 signaling pathway. Several researchers had demonstrated that the presence of p53 was necessary for DNA-damaged cells to arrest, repair the damage, and reenter the cell cycle [33]. Thus, we might suppose that those six hub genes played certain role in the progression of ccRCC and influenced the prognosis probably by regulating p53 signaling pathway, which contributed to the poor prognosis of ccRCC.

5. Conclusions

Our study used weighted gene co-expression analysis to construct a gene co-expression network, identify and validate network hub genes associated with the progression and poor prognosis of ccRCC. Eventually, six hub genes including *CCNB2*, *CDC20*, *CEP55*, *KIF20A*, *TOP2A*, and *UBE2C* were identified and validated in association with the progression and poor prognosis of ccRCC probably by regulating p53 signaling pathway.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gdata.2017.10.006.

Acknowledgments

The excellent technical assistance of Shanshan Zhang and Danni Shan is gratefully acknowledged. We also would like to acknowledge the KEGG database developed by Kanehisa Laboratories. This study was supported in part by grants from the Hubei Province Health and Family Planning Scientific Research Project (grant number WJ2017H0002 to Yu Xiao), Wuhan Clinical Cancer Research Center of Urology and Male Reproduction (grant number 303-230100055 to Xinghuan Wang) and Hubei Province Urological Clinical Research Center of Laparoscopy (grant number ZN-31 to Xinghuan Wang). The funders had no role in

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declaration of conflicts of interest

The authors declare no conflicts of interest.

References

[1] K. Gupta, J.D. Miller, J.Z. Li, M.W. Russell, C. Charbonneau, Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review, Cancer Treat. Rev. 34 (2008) 193–205.

[2] M.M. Baldewijns, I.J. van Vlodrop, L.J. Schouten, P.M. Soetekouw, A.P. de Bruine, M. van Engeland, Genetics and epigenetics of renal cell cancer, Biochim. Biophys. Acta 1785 (2008) 133–155.

[3] F.E. Vera-Badillo, A.J. Templeton, I. Duran, A. Ocana, P. de Gouveia, P. Aneja, J.J. Knox, I.F. Tannock, B. Escudier, E. Amir, Systemic therapy for non-clear cell renal cell carcinomas: a systematic review and meta-analysis, Eur. Urol. 67 (2015) 740–749.

[4] J.Y. Chan, Y. Choudhury, M.H. Tan, Predictive molecular biomarkers to guide clinical decision making in kidney cancer: current progress and future challenges, Expert. Rev. Mol. Diagn. 15 (2015) 631–646.

[5] S. Tavazoie, J.D. Hughes, M.J. Campbell, R.J. Cho, G.M. Church, Systematic determination of genetic network architecture, Nat. Genet. 22 (1999) 281–285.

[6] M.R. Stratton, P.J. Campbell, P.A. Futreal, The cancer genome, Nature 458 (2009) 719–724.

[7] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.

[8] K. Shi, Z.T. Bing, G.Q. Cao, L. Guo, Y.N. Cao, H.O. Jiang, M.X. Zhang, Identify the signature genes for diagnose of uveal melanoma by weight gene co-expression network analysis, Int. J. Ophthalmol. 8 (2015) 269–274.

[9] X.L. Zhu, Z.H. Ai, J. Wang, Xu YL, Y.C. Teng, Weighted gene co-expression network analysis in identification of endometrial cancer prognosis markers, Asian Pac. J. Cancer Prev. 13 (2012) 4607–4611.

[10] Q.L. Wang, X. Chen, M.H. Zhang, Q.H. Shen, Z.M. Qin, Identification of hub genes and pathways associated with retinoblastoma based on co-expression network analysis, Genet. Mol. Res. 14 (2015) 16151–16161.

[11] Z. Tang, C. Li, B. Kang, G. Gao, C. Li, Z. Zhang, GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses, Nucleic Acids Res. (2017), http://dx.doi.org/10.1093/nar/gkx247 [Epub ahead of print].

[12] J. Anaya, Oncolnc: linking tcga survival data to mrnas, mirnas, and lncrnas, PeerJ Computer. Sci. 2 (2016) e67.

[13] G. Dennis Jr., B.T. Sherman, D.A. Hosack, J. Yang, W. Gao, H.C. Lane, R.A. Lempicki, DAVID: database for annotation, visualization, and integrated discovery, Genome Biol. 4 (2003) P3.

[14] M.M. Croken, W. Qiu, M.W. White, K. Kim, Gene set enrichment analysis (GSEA) of toxoplasma gondii expression datasets links cell cycle progression and the bradyzoite developmental program, BMC Genomics 15 (2014) 515.

[15] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.

[16] S. Takashima, H. Saito, N. Takahashi, K. Imai, S. Kudo, M. Atari, Y. Saito, S. Motoyama, Y. Minamiya, Strong expression of cyclin B2 mRNA correlates with a poor prognosis in patients with non-small cell lung cancer, Tumour Biol. 35 (2014) 4257–4265.

[17] E. Shubbar, A. Kovacs, S. Hajizadeh, T.Z. Parris, S. Nemes, K. Gunnarsdottir, Z. Einbeigi, P. Karlsson, K. Helou, Elevated cyclin B2 expression in invasive breast carcinoma is associated with unfavorable clinical outcome, BMC Cancer 13 (2013) 1.

[18] H. Karra, H. Repo, I. Ahonen, E. Loyttyniemi, R. Pitkanen, M. Lintunen, T. Kuopio, M. Soderstrom, P. Kronqvist, Cdc20 and securin overexpression predict short-term breast cancer survival, Br. J. Cancer 110 (2014) 2905–2913.

[19] Z.Y. Ding, Wu HR, J.M. Zhang, G.R. Huang, D.D. Ji, Expression characteristics of CDC20 in gastric cancer and its correlation with poor prognosis, Int. J. Clin. Exp. Pathol. 7 (2014) 722–727.

[20] Wu WJ, Hu KS, D.S. Wang, Z.L. Zeng, D.S. Zhang, D.L. Chen, L. Bai, R.H. Xu, CDC20 overexpression predicts a poor prognosis for patients with colorectal cancer, J. Transl. Med. 11 (2013) 142.

[21] Z. Wang, L. Wan, J. Zhong, H. Inuzuka, P. Liu, F.H. Sarkar, W. Wei, Cdc20: a potential novel therapeutic target for cancer treatment, Curr. Pharm. Des. 19 (2013) 3210–3214.

[22] J. Tao, X. Zhi, Y. Tian, Z. Li, Y. Zhu, W. Wang, K. Xie, J. Tang, X. Zhang, L. Wang, Z. Xu, CEP55 contributes to human gastric carcinoma by regulating cell proliferation, Tumour Biol. 35 (2014) 4389–4399.

[23] J. Jeffery, D. Sinha, S. Srihari, M. Kalimutho, K.K. Khanna, Beyond cytokinesis: the emerging roles of CEP55 in tumorigenesis, Oncogene 35 (2016) 683–690.

[24] C. Shi, D. Huang, N. Lu, D. Chen, M. Zhang, Y. Yan, L. Deng, Q. Lu, H. Lu, S. Luo, Aberrantly activated Gli2-KIF20A axis is crucial for growth of hepatocellular carcinoma and predicts poor prognosis, Oncotarget 7 (2016) 26206–26219.

[25] W. Zhang, W. He, Y. Shi, H. Gu, M. Li, Z. Liu, Y. Feng, N. Zheng, C. Xie, Y. Zhang, High expression of KIF20A is associated with poor overall survival and tumor progression in early-stage cervical squamous cell carcinoma, PLoS One 11 (2016) e0167449.

[26] J.S. Kirk, K. Schaarschuch, Z. Dalimov, E. Lasorsa, S. Ku, S. Ramakrishnan, Q. Hu, G. Azabdaftari, J. Wang, R. Pili, L. Ellis, Top2a identifies and provides epigenetic rationale for novel combination therapeutic strategies for aggressive prostate cancer, Oncotarget 6 (2015) 3136–3146.

[27] J. Lan, H.Y. Huang, S.W. Lee, T.J. Chen, H.C. Tai, H.P. Hsu, K.Y. Chang, C.F. Li, TOP2A overexpression as a poor prognostic factor in patients with nasopharyngeal carcinoma, Tumour Biol. 35 (2014) 179–187.

[28] Y.C. Xu, F.C. Zhang, J.J. Li, J.Q. Dai, Q. Liu, L. Tang, Y. Ma, Q. Xu, X.L. Lin, H.B. Fan, H.X. Wang, *RRM1*, *TUBB3*, *TOP2A*, *CYP19A1*, *CYP2D6*: difference between mRNA and protein expression in predicting prognosis of breast cancer patients, Oncol. Rep. 34 (2015) 1883–1894.

[29] Z. Zhang, P. Liu, J. Wang, T. Gong, F. Zhang, J. Ma, N. Han, Ubiquitin-conjugating enzyme E2C regulates apoptosis-dependent tumor progression of non-small cell lung cancer via ERK pathway, Med. Oncol. 32 (2015) 149.

[30] T. Morikawa, T. Kawai, H. Abe, H. Kume, Y. Homma, M. Fukayama, UBE2C is a marker of unfavorable prognosis in bladder cancer after radical cystectomy, Int. J. Clin. Exp. Pathol. 6 (2013) 1367–1374.

[31] A. Psyrri, K.T. Kalogeras, R. Kronenwett, R.M. Wirtz, A. Batistatou, E. Bournakis, E. Timotheadou, H. Gogas, G. Aravantinos, C. Christodoulou, T. Makatsoris, H. Linardou, D. Pectasides, N. Pavlidis, T. Economopoulos, G. Fountzilas, Prognostic significance of UBE2C mRNA expression in high-risk early breast cancer. A Hellenic Cooperative Oncology Group (HeCOG) study, Ann. Oncol. 23 (2012) 1422–1427.

[32] D. Loussouarn, L. Campion, F. Leclair, M. Campone, C. Charbonnel, G. Ricolleau, W. Gouraud, R. Bataille, P. Jezequel, Validation of UBE2C protein as a prognostic marker in node-positive breast cancer, Br. J. Cancer 101 (2009) 166–173.

[33] D.J. Lukin, L.A. Carvajal, W.J. Liu, L. Resnick-Silverman, J.J. Manfredi, p53 Promotes cell survival due to the reversibility of its cell-cycle checkpoints, Mol. Cancer Res. 13 (2015) 16–28.